INFECTION AND IMMUNITY, May 1992, p. 1890–1893
0019-9567/92/051890-04$02.00/0

Vol. 60, No. 5

Evidence Consistent with Horizontal Transfer of the Gene (*emm12*) Encoding Serotype M12 Protein between Group A and Group G Pathogenic Streptococci

WARREN J. SIMPSON,[1]† JAMES M. MUSSER,[2,3] AND P. PATRICK CLEARY[1]*

Department of Microbiology, University of Minnesota, Minneapolis, Minnesota 55455[1]; Institute of Molecular Evolutionary Genetics, Pennsylvania State University, University Park, Pennsylvania 16802[2]; and Department of Pathology, Baylor College of Medicine, Houston, Texas 77030[3]

Received 6 November 1991/Accepted 10 February 1992

Human isolates of Lancefield group G streptococci harbor sequences homologous with the structural gene (*emm*) encoding M protein, a major virulence factor in *Streptococcus pyogenes* (a group A *Streptococcus* species). We used DNA-DNA hybridization, restriction endonuclease chromosomal profiling, and multilocus enzyme electrophoresis to examine genetic relationships between group A and group G streptococcal strains expressing homologous serologic type 12 M (M12) protein. All M12 group A strains studied had very similar restriction endonuclease genomic profiles and multilocus enzyme genotypes. In contrast, the restriction enzyme genomic profile and multilocus enzyme genotype of the M12 group G strain CS140 were strikingly different from those characterizing the M12 group A organisms. DNA-DNA hybridization studies revealed, on average, 57% genomic similarity between the M12 group A and group G strains. Taken together, our data demonstrate that a gene encoding M12 protein occurs in two highly divergent chromosomal backgrounds, a result suggesting that an episode of horizontal gene transfer and recombination has occurred between two streptococcal lineages.

M protein is a major virulence determinant of the human pathogen *Streptococcus pyogenes* (a group A *Streptococcus* species) (7, 11). The molecule is associated with the bacterial cell surface and confers resistance to phagocytosis by polymorphonuclear leukocytes (11). Serologic studies have revealed the existence of at least 80 distinct M protein antigenic types (7). Although a strong immune response is elicited by a streptococcal infection, opsonic antibodies are protective only against strains expressing the homologous M protein serotype. As a consequence of the large variety of M protein serotypes and the restricted production of opsonic antibodies to the homologous serotype, the risk of multiple infection episodes throughout life is high. The molecular mechanisms responsible for generating allelic variations in the gene encoding the M protein antigen are poorly understood, although intragenic rearrangements can occur (8).

Many streptococcal isolates expressing the Lancefield group G antigen recovered from human infections harbor sequences hybridizing with probes derived from the gene encoding the M protein in group A organisms (17). Group G organisms recovered from diseased humans also display immunologic heterogeneity in M proteins (1, 9, 12, 13). Interestingly, group G isolates cultured from animal infections fail to hybridize with *emm* gene probes, an observation that has been interpreted as evidence supporting the notion that M protein expression is an important virulence factor for humans (17).

The molecular explanation for the presence of an M protein gene in virtually all Lancefield group A isolates and in all human pathogenic group G isolates studied but its absence in animal group G disease isolates is unknown. A variety of hypotheses can account for these observations. For example, it is possible that a restricted subset of group G isolates has acquired an M protein gene by mutational convergence. An alternative hypothesis is that horizontal transfer and recombination events have moved an M protein structural gene from a Lancefield group A organism to an isolate expressing the group G antigen (recombinational convergence). A third hypothesis is that a lineage expressing the M protein and the group A antigen has experienced genomic changes, by either recombinational or mutational convergence, resulting in the expression of a group G carbohydrate antigen. Because little is known about the levels of genetic diversity among and between isolates assigned to group A and group G, it is not possible to differentiate among the various hypotheses. We used hybridization of whole genomic DNA, multilocus enzyme electrophoresis, and restriction endonuclease profiling of chromosomal DNA (genomic fingerprinting) to generate molecular data concerning the mechanism responsible for the presence of an *emm* gene in isolates of two distinct Lancefield serogroups.

MATERIALS AND METHODS

Bacterial isolates. Strain CS140, recovered from a diseased human, is a group G isolate that expresses type 12 M (M12) antigen (3, 17). Strains CS24, 9421, and 87-244 were cultured from epidemiologically unrelated human infections, express M12 antigen, but are classified as Lancefield group A (17). Additional group A and group G isolates were obtained from the World Health Organization Collaborating Center for Reference and Research on Streptococci, located at the University of Minnesota. Some of the isolates were partially characterized previously (17).

DNA restriction endonuclease profiling and DNA-DNA hybridization. Media, growth conditions, and DNA preparation protocols were described previously (17–19). We used previously formulated protocols for DNA fingerprinting (4, 18, 19) and DNA-DNA hybridization (10). Labelled genomic DNA and DNA filters were prepared as described by Taylor

* Corresponding author.
† Present address: North East Valley, Dunedin, New Zealand.

et al. (20). In brief, total genomic DNA was denatured, bound to nitrocellulose filters, and hybridized with a ^{32}P-labelled probe made by nick translation (nick translation kit; Bethesda Research Laboratories, Gaithersburg, Md.) of genomic DNA purified from a second strain.

Multilocus enzyme electrophoresis. Methods of protein extract preparation, electrophoresis of soluble enzymes, and selective enzyme staining were described previously (14–16). The eight enzymes assayed were adenylate kinase, carbamylate kinase, lactate dehydrogenase, phosphoglucose isomerase, nucleoside phosphorylase, leucylglycylglycine peptidase, esterase, and β-glucuronidase. Distinctive electromorphs (mobility variants) of each enzyme, numbered in order of decreasing rate of anodal migration, were equated with alleles at the corresponding structural gene locus (16). Isolates that lacked activity for a specific enzyme were assigned a null allelic state at the locus in question. Each isolate was characterized by its combination of alleles at the eight enzyme loci, and distinctive multilocus enzyme genotypes were designated electrophoretic types (16).

RESULTS AND DISCUSSION

DNA restriction endonuclease profiling. The results of restriction endonuclease genomic profiling of a variety of epidemiologically unassociated strains after digestion with *Hin*dIII and *Hae*III are presented in Fig. 1 and 2, respectively. Among the strains expressing M12 protein and group A carbohydrate, no variation was detected after cleavage with *Hin*dIII (Fig. 1) and only nominal restriction fragment heterogeneity was revealed after digestion with *Hae*III (Fig. 2). In contrast, the genomic *Hin*dIII and *Hae*III profiles generated for strain CS140 (expressing M12 and group G antigens) were strikingly distinctive from those of all M12 group A strains. The *Hae*III restriction profile of CS140 also was different from that pattern for six additional epidemiologically unrelated group G strains cultured from infected humans (Fig. 2). Taken together, the data suggest that in overall genomic organization, group G strain CS140 is significantly divergent from group A strains expressing the same M protein serotype. The data also demonstrate a relative lack of genomic heterogeneity among M12 group A human isolates compared with group G human isolates. Our data also demonstrate that, with regard to other group G isolates causing human infections, strain CS140 is relatively distinct in chromosomal character.

DNA-DNA hybridization. The results of the filter hybridization experiments are presented in Table 1. These experiments revealed that strain CS140 shares less than 60% genomic DNA homology with strains CS24 and 9421, two representative group A strains expressing M12 protein. In contrast, DNA hybridization among the three independent group A M12 isolates was significantly higher, being, on average, about 80%. These results demonstrate that a gene encoding M12 protein is harbored by organisms that are highly divergent in overall genomic character. Hence, strain CS140 has not shared a recent ancestor with any of the group A strains examined here. The low (<60%) hybridization that we observed agrees with the results of other studies demonstrating that some group A and group G (large-colony-forming type) organisms represent separate species (5, 6, 10). Inasmuch as the commonly accepted species level hybridization is 70% (2, 22), our results are consistent with the notion that M12 protein is expressed by strains that are specifically distinct. Therefore, the data effectively rule out the possibility of a recent interconversion of an M12-expressing group A strain to a group G serotype, or vice versa, at least among the small sample of genotypes that we studied.

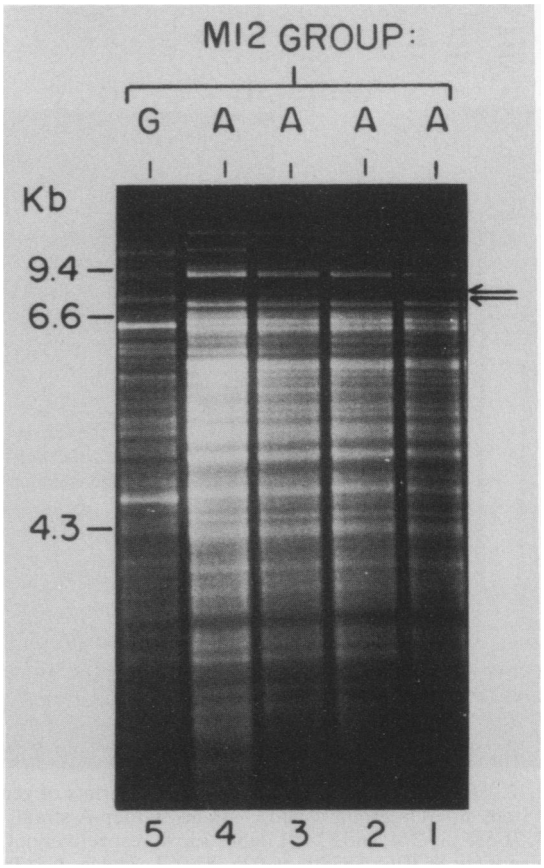

FIG. 1. *Hin*dIII restriction endonuclease fingerprints of genomic DNAs from different M12 group A and G isolates. Group A strains were CS24 (lane 1), 86-494 (Utah) (lane 2), 87-113 (Argentina) (lane 3), and 87-244 (South Dakota) (lane 4). Group G strain CS140 (Trinidad) is in lane 5 (17). The two arrows denote fragments unique to strains CS24 and 87-113. Molecular weight markers were *Hin*dIII-digested lambda DNA fragments. Digested DNA was electrophoresed on an 0.8% gel at 40 V for 15 h.

Multilocus enzyme electrophoresis. The results of multilocus enzyme electrophoretic analysis are presented in Table 2. Among the five M12 group A isolates studied, two electrophoretic types were identified. Strains CS24, 86-494, 87-091, and 86-766 had identical allele profiles, and strain 86-375 had a single-locus variant genotype (possession of the "10" allele for the locus encoding phosphoglucose isomerase). These data are consistent with the results generated from restriction endonuclease profiling, which demonstrated a relative lack of genomic heterogeneity among this small sample of M12 strains.

Significantly more genomic heterogeneity was present among the five isolates classified as Lancefield group G. Our analysis revealed that each isolate, including CS140, had a distinctive multilocus enzyme genotype and that certain group G isolates (for example, strains 76-428 and CS138) shared no particularly close overall genomic relationship. Moreover, there was very little sharing of alleles between group A and group G strains, a result fully consistent with the significant levels of genomic differentiation revealed by DNA-DNA hybridization.

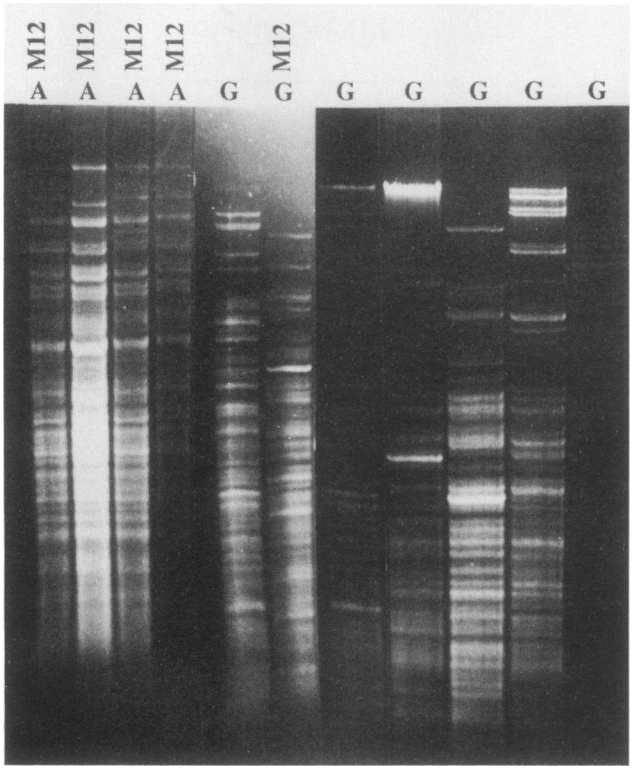

FIG. 2. *Hae*III restriction endonuclease fingerprints of genomic DNAs from different group A and G strains. Group A strains were 86-494, 71-685, 87-244, and 87-091 (lanes 1 to 4, respectively). Group G strains were 86-026, CS140, 86-026, 87-057, 74-446, CS138, and 76-428 (lanes 5 to 11, respectively).

Our experiments establish that group G strain CS140 is not a recently evolved variant of a common M12 group A streptococcal genotype. The data suggest that a horizontal gene transfer and recombination event involving movement of part or all of an *emm* gene from an M12 group A donor to a group G recipient has occurred. In this regard, it may be noteworthy that Totolian (21) has reported that M protein genes can be horizontally transferred among group A strains by bacteriophage transduction.

In theory, our data could be explained by mutational convergence. However, the demonstration that a 20-mer oligonucleotide probe corresponding to a common region in the signal sequences of structural genes (*emm1*, *emm6*, *emm12*, and *emm24*) encoding other group A M protein antigens hybridizes to the group G *emm12* gene under stringent conditions allowing only a single base-pair mismatch (17) argues against a hypothesis invoking convergence by mutation.

One additional hypothesis for which we currently lack data either to support or to reject is that strain CS140 has arisen from an unidentified genotype encoding either the M12 protein antigen or a related antigen. In this hypothesis, an ancestral cell would have acquired the ability to express the group G antigen either by mutational convergence or by lateral gene flow. The hypothesis suggests that there may be human group A streptococcal strains more closely related in overall chromosomal character to human group G streptococcal strains than the M12 genotypes that we examined in the present study. Studies are under way to test this idea.

TABLE 1. DNA similarities among various strains

Filter-bound DNA (strain)	% Relative binding[a] of labelled DNA from strain:		
	CS24	9421	CS140
CS24 (group A, M12)	100	84	56
9421 (group A, M12)	89	100	50
87-244 (group A, M12)	100	100	ND
CS140 (group G, M12)	51	59	100
Bacillus sp.	6	ND	4
Staphylococcus sp.	10	ND	8

[a] Percent relative binding is expressed as the percentage of labelled DNA bound to filters with unlabeled fixed DNA, relative to that bound to filters with homologous DNA. Hybridization was carried out at melting temperature −25°C. All tests were performed in triplicate. ND, not done.

TABLE 2. Allele profiles for group G and A streptococci

Group	Strain[a]	Allele profile for the following enzyme locus[b]:							
		AK	CAK	LDH	PGI	NSP	LGG	EST	bGLU
G	CS140	5/7	0	3	4	2	5/10	5	7
	74-446	5/7	0	0	8	3	5/8	5	5
	CS138	5/7	0	3	6	3	5/10	5	7
	86-057	7	0	3	4	2	5/10	7	7
	86-026	7	7	0	4	5	7	0	7
A	CS24	3	0	5	6	5	8	1	0
	86-494	3	0	5	6	5	8	1	0
	87-091	3	0	5	6	5	8	1	0
	86-766	3	0	5	6	5	8	1	0
	86-375	3	0	5	10	5	8	1	0

[a] Strain 86-026 was isolated from a dog. The other group G strains were of human origin.

[b] AK, adenylate kinase; CAK, carbamylate kinase; LDH, lactose dehydrogenase; PGI, phosphoglucose isomerase; NSP, nucleoside phosphorylase; LGG, leucylglycylglycine peptidase; EST, esterase; bGLU, β-glucuronidase. A slash indicates more than one allele.

ACKNOWLEDGMENTS

We thank J. Smith for secretarial assistance.

W.J.S. was supported by a fellowship from the American Heart Association, Minnesota Affiliate, Inc. This research was supported by grant AI-16722 (to P.P.C.) from the National Institutes of Health and by biomedical research support grant RR-05425 (to J.M.M.) from the National Institutes of Health.

REFERENCES

1. **Bisno, A. L., D. E. Craven, and W. R. McCabe.** 1987. M proteins of group G streptococci isolated from bacteremic human infections. Infect. Immun. **55:**753–757.
2. **Brenner, D. J.** 1984. Family 1. *Enterobacteriaceae* Rahn 1937, p. 408–420. *In* N. R. Krieg and J. G. Holt (ed.), Bergey's manual of systematic bacteriology, vol. 1. The Williams & Wilkins Co., Baltimore.
3. **Cleary, P., J. Peterson, C. Chen, and C. Nelson.** 1991. Virulent human strains of group G streptococci express a C5a peptidase enzyme similar to that produced by group A streptococci. Infect. Immun. **59:**2305–2310.
4. **Cleary, P. P., E. L. Kaplan, C. Livdahl, and S. Skjold.** 1988. DNA fingerprints of *Streptococcus pyogenes* are M type specific. J. Infect. Dis. **158:**1317–1323.
5. **Devriese, L. A., J. Hommez, R. Kilpper-Balz, and K. H. Schleifer.** 1986. *Streptococcus canis* sp. nov.: a species of group G streptococci from animals. Int. J. Syst. Bacteriol. **36:**422–425.
6. **Farrow, J. A. E., and M. D. Collins.** 1984. Taxonomic studies on streptococci of serological groups C, G and L and possibly related taxa. Syst. Appl. Microbiol. **5:**483–493.

7. **Fischetti, V. A.** 1989. Streptococcal M protein: molecular design and biological behavior. Clin. Microbiol. Rev. **2**:285–314.
8. **Hollingshead, S. K., V. A. Fischetti, and J. R. Scott.** 1987. Size variation in group A streptococcal M protein is generated by homologous recombination between intragenic repeats. Mol. Gen. Genet. **207**:196–203.
9. **Jones, K. F., and V. A. Fischetti.** 1987. Biological and immunochemical identity of M protein on group G streptococci with M protein on group A streptococci. Infect. Immun. **55**:502–506.
10. **Kilpper, R., U. Buhl, and K. H. Schleifer.** 1980. Nucleic acid homology studies between *Peptococcus saccharolyticus* and various anaerobic and facultative anaerobic gram-positive cocci. FEMS Microbiol. Lett. **8**:205–210.
11. **Lancefield, R. C.** 1962. Current knowledge of type-specific M antigens of group A streptococci. J. Immunol. **89**:307–313.
12. **Lawal, S. F., A. O. Coker, E. O. Solanke, and O. Ogunbi.** 1982. Serotypes among Lancefield-group G streptococci isolated in Nigeria. J. Med. Microbiol. **15**:123–125.
13. **Maxted, W. R., and E. V. Potter.** 1967. The presence of type 12 M-protein antigen in group G streptococci. J. Med. Microbiol. **49**:119–125.
14. **Musser, J. M., A. R. Hauser, M. H. Kim, P. M. Schlievert, K. Nelson, and R. K. Selander.** 1991. *Streptococcus pyogenes* causing toxic-shock-like syndrome and other invasive diseases: clonal diversity and pyrogenic exotoxin expression. Proc. Natl. Acad. Sci. USA **88**:2668–2672.
15. **Musser, J. M., S. J. Nattingly, R. Quentin, A. Goudeau, and R. K. Selander.** 1989. Identification of a high-virulence clone of type III *Streptococcus agalactiae* (group B streptococcus) causing invasive neonatal disease. Proc. Natl. Acad. Sci. USA **86**:4731–4735.
16. **Selander, R. K., D. A. Caugant, H. Ochman, J. M. Musser, M. N. Gilmour, and T. S. Whittam.** 1986. Methods of multilocus enzyme electrophoresis for bacterial population genetics and systematics. Appl. Environ. Microbiol. **51**:873–884.
17. **Simpson, W. J., J. C. Robbins, and P. P. Cleary.** 1987. Evidence for group A-related M protein genes in human but not animal-associated group G streptococcal strains. Microb. Pathog. **3**:339–350.
18. **Skjold, S. A., and P. P. Cleary.** 1987. DNA fingerprints of group A, serotype M49 streptococci, p. 127–129. *In* J. J. Ferretti and R. Curtiss III (ed.), Streptococcal genetics. American Society for Microbiology, Washington, D.C.
19. **Skjold, S. A., P. G. Quie, L. A. Fries, M. Barnham, and P. P. Cleary.** 1987. DNA fingerprinting of *Streptococcus zooepidemicus* (Lancefield group C) as an aid to epidemiological study. J. Infect. Dis. **155**:1145–1150.
20. **Taylor, A. J., C. A. Dawson, and R. J. Owen.** 1986. The identification of *Bacteroides ureolyticus* from patients with non-gonococcal urethritis by conventional biochemical tests and by DNA and protein analyses. J. Med. Microbiol. **21**:109–116.
21. **Totolian, A.** 1979. Transduction of M-protein and serum-opacity-factor production in group A streptococci, p. 38–39. *In* M. T. Parker (ed.), Pathogenic streptococci. Reedbooks Ltd., Windsor, England.
22. **Wayne, L. G., D. J. Brenner, R. R. Colwell, P. A. D. Grimont, O. Kandler, M. I. Krichevsky, L. H. Moore, W. E. C. Moore, R. G. E. Murray, E. Stackebrandt, M. P. Starr, and H. G. Trüper.** 1987. Report of the ad hoc committee on reconciliation of approaches to bacterial systematics. Int. J. Syst. Bacteriol. **37**:463–464.